Multimodal Biometric Authentication using Convolutional Neural Network, Swin Transformer, and Multi-Head Attention with Global Max Pooling

Shafiq Sharhrah[1], *Divya.* G[2], *Hareesh.* K N[3*], *Babu* Rajendra Prasad[4], and *K. P.* Anusuya[5]

[1]Department of Computers Techniques Engineering, College of Technical Engineering, The Islamic University, Najaf, Iraq
[2]Department of Electronics and Communication Engineering, Nitte Meenakshi Institute of Technology, Nitte (Deemed to be University), Bengaluru, India
[3]Department of Electronics and Telecommunication Engineering, Sri Siddhartha Institute of Technology, Tumkur, India
[4]Vidyavardhaka College of Engineering, VTU, Mysuru, India
[5]Department of Humanities, Dayananda Sagar College of Engineering, Bengaluru, India

Abstract. Multimodal biometric authentication is a robust security mechanism designed to enhance the reliability and security of user-authentication systems. It integrates several biometric traits to provide an accurate and secure identification system. However, existing Deep Learning (DL) models struggle to capture both spatial and contextual dependencies across multiple biometric traits, which reduces their robustness against spoofing attacks. Hence, this paper proposes a convolutional neural network, swin transformer, multi-head self-attention, and global max pooling (CSMG) for effective feature extraction, strengthening both spatial and contextual representation, and reducing redundancy across modalities. Next, the classification head uses the extracted feature map and softmax layer to predict a person's identity. An effective fusion strategy was introduced to integrate fingerprint, iris, and ECG signals, utilizing their complementary strengths to mitigate spoofing attacks. The performance of the proposed CSMG method was evaluated using the IITD Iris, SOCOfing fingerprint, and HEARTPRINT ECG datasets. The experimental evaluation demonstrates that the proposed CSMG method achieves a recognition accuracy of 99.90% for fingerprints, 99% for irises, and 99% for ECG compared to traditional models.

1 Introduction

In recent years, biometric authentication has gained wider use than the standard usernames and password authentication [1]. The security of personal and organizational data has risen to the forefront of technological innovation and research with the advent of an increasingly digital and interconnected world [2]. Security, access control, and forensic investigations rely on biometric systems that use physiological or behavioral traits to identify individuals [3]. There are two types of biometric methods: unimodal and multimodal. Unimodal approaches use only one biometric trait for identification, frequently resulting in low accuracy and susceptibility to spoofing [4]. Although a single biometric method is effective, several studies have demonstrated that combining multiple systems yields better results [5]. In multimodal systems, the fusion of processed data can take place at different levels depending on the modality, with decision and score levels commonly applied across all modalities [6]. Moreover, integrating multi-biometric systems that combine several biometric traits offers the potential to improve recognition accuracy and robustness [7]. multi-biometric fusion has gained significant attention because it offers reliable authentication and enhances system security [8]. In multimodal biometric recognition, combining multiple features has become standard practice over using a single biometric feature, thereby enhancing biometric security systems [9]. Multimodal biometric systems help safeguard both stored and transmitted data, mitigating the risks of unauthorized access and data breach [10].

The key contribution of this paper describe below:

1. Propose a Deep Learning (DL) architecture (CSMG-BioNet) that combines a CNN, Swin Transformer, Multi-Head Self-Attention, and Global Max Pooling to effectively extract features from multiple biometric modalities.
2. Integrates fingerprint, iris, and ECG signals using an effective multimodal fusion strategy, leveraging complementary strengths to enhance the recognition accuracy and mitigate spoofing attacks.
3. Strengthens both spatial and contextual representations of biometric features while reducing redundancy across modalities and improving robustness compared to traditional deep learning models.
4. A classification head is employed with a softmax layer that utilizes the extracted feature

* Corresponding author: hareeshkn@ssit.edu.in

map to accurately predict an individual's identity across multiple biometric traits.

The following description of remaining sections, Section 2 provides a comprehensive literature review, Section 3 outlines the methodologies, Section 4 presents the results and comparative analysis of the Multimodal Biometric Authentication and Section 5 discusses the overall findings, and Section 6 concludes with future work.

2 Literature Review

This section discusses some existing DL and ML models based Multimodal biometric authentication system.

Reena Garg et al. [11] developed a feature extraction method in which attention-based and spatial features are extracted in parallel. Fusion of fingerprint, iris, and ECG was used with a convolutional Swim Transformer for effective feature extraction, exploitation of complementary strengths, and mitigation of spoofing risks. For the evaluation, multimodal datasets were used to achieve a better recognition accuracy. The developed method advanced multimodal biometric systems by demonstrated the effectiveness of the robust fusion and feature extraction methods. However, multimodal biometric systems still lack complementary strength and feature extraction capability, which reduces accuracy and robustness.

M. Ramkumar Prabhua et al. [12] developed a secure multi-biometric system that relied on feature-level fusion for user identification. Fingerprint images were normalized using ResNet50 extracted high-level features from ECG signals and fingerprints, while Min-Max reduced variations in image quality. The developed method was implemented in Python, and the experimental results demonstrated that the developed multi-biometric method achieved higher performance and proved to be dependable, robust, and effective than existing authentication methods. Despite this improvement, challenges remained in enhanced scalability and sensor speed to ensure applicability in diverse scenarios and continued relevance in a rapidly evolving technological environment.

Farhad Ahamed et al. [13] Employed an AI-based multimodal biometric authentication method for single and group-based device-level authentication using ECG and PPG signals. The employed method integrated the physiological features from multiple datasets into a unified authentication method, enabling continuous monitoring and collaborative validation without relying on a single biometric modality. Fusion techniques have been applied to enhance authentication accuracy and strengthen security. The employed method demonstrated strong performance across five public datasets from PhysioNet and Mendeley Data, achieving higher accuracy and an Equal Error Rate (EER). However, the employed method faced challenges in handling noisy or low-quality signals, which affected the model reliability and increased the computational requirements for real-time deployment.

Jananee Vinayagam et al. [14] developed an effective multimodal BA method used a Kernel Correlation Padding-based Deep CNN (KCP-DCNN). Input images preprocessed for magnification via RBF-PRT and augmented using LZS-GAN were combined with signature, fingerprint, and face modalities. FDAF-SACM contour extraction, chaincode-based minutia extraction, and Dlib's 68 facial point extraction were performed on the processed images. Key features were then extracted using the KCP-DCNN for recognition, followed by verification through QR codes. Evaluation using the SOCOFing and Face recognition datasets demonstrated that the developed method achieved better performance and more effective recognition. However, extensive preprocessing and multiple feature extraction steps are required, which increases the computational complexity and makes the implementation more challenging.

Rajneesh Rani et al. [15] developed an effective cancelable iris biometric method utilizing a hybrid CNN and Gated Recurrent Unit (GRU) method and used a random projection technique to cancel the system. Iris images from the IIT Delhi (IITD) and MMU datasets were preprocessed for normalization and enhancement, and significant iris features were extracted using the CNN–GRU hybrid method for robust recognition. The method was validated using an Equal Error Rate (EER), and improved identification accuracy and strong resistance against template compromise were achieved. However, the method required complex hybrid method training, and the increased computational overhead made the implementation more resource-intensive than conventional iris recognition approaches.

2.1 Problem

Accurate biometric authentication is challenging due to the limitations of traditional DL methods in capturing both spatial and contextual dependencies across multiple biometric traits. Single-modality systems are vulnerable to spoofing attacks, and conventional methods often fail to exploit the complementary information provided by fingerprints, iris, and ECG signals, resulting in reduced recognition accuracy and classification.

2.2 Objective

To address these challenges, this paper proposes a CSMG method that combines Convolutional Neural Networks, Swin Transformers, Multi-Head self-attention, and global max pooling to effectively extract hierarchical and contextual features from multiple biometric modalities. An effective multimodal fusion strategy integrates fingerprint, iris, and ECG signals, improving feature images, reducing redundancy, and enhancing recognition accuracy while mitigating spoofing attacks across diverse datasets.

3 Proposed Method

This section describes the structure and components of the proposed CSMG method. This paper proposes a CSMG method that combines CNN, Swin Transformer,

MHSA, and Global Max Pooling for effective feature extraction, strengthens both spatial and contextual representations, and reduces redundancy across modalities. The CSMG method works with multimodal datasets: IITD Iris, SOCOfing fingerprint, and HEARTPRINT ECG. The preprocessed and input to the CNN learns the edges and textures. The Swin Transformer is used to capture wider and more detailed relationships in the data. MHSA highlights the most important features by providing a key part of the biometric patterns. Subsequently, the Global Max Pooling layer reduces the feature size and retains the important information. Finally, these features are passed through a fully connected layer with a softmax function to identify the person. The CSMG method improves accuracy and classification. The overall architecture of proposed CSMG method is shown in Figure 1.

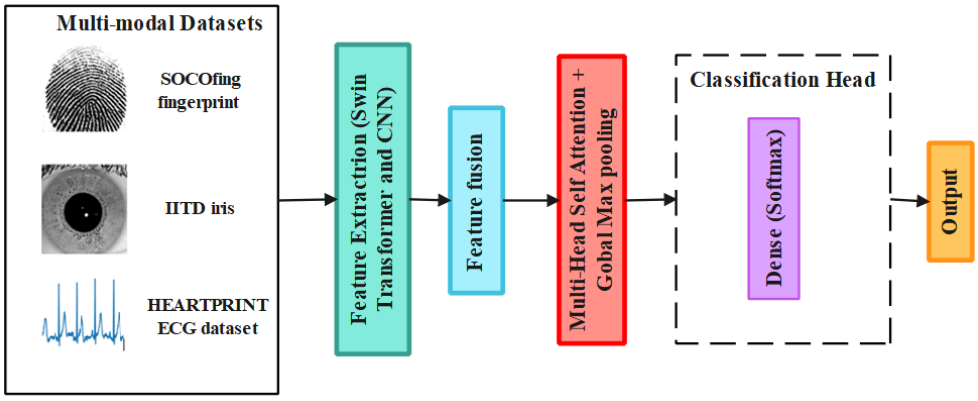

Fig. 1. Overall architecture of proposed CSMG method.

3.1 Dataset Descriptions

In this section, multimodal biometric classification uses multimodal datasets such as fingerprint, iris, and ECG data.

The fingerprint dataset used was SOCOFing [16], contains 6,000 images, 10 per person, from 600 individuals. The images were center-cropped, resized, and reshaped for uniformity. Iris images were obtained from the IITD dataset [17], consists of 2,240 images, 10 per person, from 224 individuals. The region of interest in each iris image was localized using the Hough Transform and then normalized to ensure consistency across all images. ECG [18] recordings were taken from Heartprint dataset, which includes data from 199 participants aged 18–68 years, mainly from South Asian and Arab backgrounds. Captured from fingertips, signals used dry electrodes integrated into a ReadMyHeart device. Each participant had minimum two 15-second recordings per session, sampled at 250 Hz, with multiple sessions collected over time.

3.2 Feature Extraction Network

The Feature Extraction Network combines CNN method and Swin Transformer as proposed.

3.2.1 Swin Transformer (ST)

In this work, the ST is used solely for feature extraction, and each biometric image is resized before feature extraction. The input image is split into patches for processing by ST. Unlike ViT, which employs fixed-size patches, ST progressively increases the patch size in deeper layers. This hierarchical approach captures by merging patches in later layers using a shifted-window partitioning strategy, detailed feature maps are created. In the current layer, self-attention is calculated within each window, and in the subsequent layer, the window is shifted to establish inter-window connections. The hierarchical structure of ST allows processing features at multiple scales while keeping computational demands proportional to the image size. This capability makes it well-suited for various computer vision tasks, with each component contributing to the extraction of features from biometric images.

3.2.1.1 Patch partitioning

The input image, divided into non-overlapping patches, has each patch forming a small feature map. These patches considered tokens and were processed sequentially through successive blocks of the ST. This method enables the capture of both local and global dependencies, preserving spatial information while progressively learning more abstract and hierarchical feature representations for each biometric modality.

3.2.1.2 Swin transformer blocks with four stages

First, the raw feature vector is projected onto a higher-dimensional space using a linear embedding layer, and the resulting patches or tokens are then processed through multiple ST blocks. This process maintains a consistent token structure throughout the network, allowing the model to effectively capture hierarchical and spatial relationships across feature representations for enhanced feature extraction.

3.2.2 Convolutional Neural Network model

CNN, a commonly used DL model in computer vision, effectively identifies hidden patterns and features within images and is suitable for classification tasks. Each

image is resized and passed through the CNN model, structured into four blocks, each containing a maxpooling layer, BN layer, and convolution layer.

3.2.2.1 Convolution Layer (CL)

Each CL transforms low-level image features into high-level representations by convolving input feature map with learnable filters generate an output feature map. For each layer, input x is a 3D array, where 2D represents the size of the input feature maps and 3D array, with 2D representing the spatial size of output feature maps and 3D corresponding to number of filters applied in convolutional layer. This process allows each convolution layer to progressively extract more complex and abstract features from the input data, thereby enabling network to capture hierarchical patterns and spatial dependencies. The output feature map y_j is expressed in Equation (1).

$$y_j = (F \otimes x_i + b_j) \quad (1)$$

Here, $\otimes$ denotes the convolution process using a 2D filter F and x_i where b_j is the bias.

3.2.2.2 Batch normalization layer (BN)

The BN layer normalizes and regularizes each convolutional layer's output stabilizing training, reduces overfitting, and enhances CNN training speed.

3.2.2.3 Maxpooling layer

Each convolution, max pooling layer to reduce feature map dimensionality and retain key features, as expressed in Equation (2):

$$Y_j = max\ (x_{ik} \forall k : 1 \leq k \leq m * m) \quad (2)$$

Hence, it outputs feature map y_j by selecting maximum value from the input feature map x_{ik} of size mXm, which is part of x_i.

3.2.3 Feature fusion

Feature vectors obtained from the ST and CNN models are concatenated to form a single combined feature map. The top features are individually selected for each modality from this integrated map. A scoring function in the Python scikit-learn library measured the statistical dependency between each feature and its corresponding variable, with mutual information used in selection process. Features with higher information values were considered to be more relevant for the prediction task. Feature sets from all modalities are combined after selection to create a unified feature map. This combined map is then further refined to extract the most salient subset of features, ensuring that the final representation retained the most informative and discriminative elements necessary for accurate classification and recognition tasks.

3.2.4 Multi-Head Self-Attention Mechanism (MHSA) with Global pooling

MHSA builds upon the self-attention mechanism by using multiple attention heads to capture more information from the input. This improves the model's understanding of the feature information by capturing diverse associations and dependencies within the features. The features extracted by an autoencoder contain some redundancy as well as irrelevant and hidden information. In this paper, MHSA used to capture better internal correlations among features at different positions. The MHSA mechanism allows for the identification of key features crucial for predicting events using critical information from omics and individual omics data. Global Max Pooling (GMP) selects the maximum value from each feature map, effectively retaining the most significant features while discarding less important information. This mechanism emphasizes key feature activations, improves the model's ability to capture critical patterns, and reduces redundancy in extracted representations. The formula for computational multi-head attention is as follows Equations (3)-(6):

$$Head_i = softmax\left(\frac{Q_i \times K_i^T}{\sqrt{d_k}}\right) V_i \quad (3)$$

$$Q_i = H \times W_i^Q \quad (4)$$

$$K_i = H \times W_i^K \quad (5)$$

$$V_i = H \times W_i^V \quad (6)$$

$Q_i, K_i, and\ V_i$ are obtained from the linear transformations of the latent vector H, with $W_i^Q, W_i^K, and\ W_i^V$ representing the parameter matrix.

3.2.5 Classification Head using Softmax Layer

In the classification head, a softmax activation function is used for the final recognition, with each output node representing an individual class. The feature layer map obtained from the previous layers is fed into the softmax layer, which calculated the probability of the input across all classes. For each biometric sample, the probabilities are calculated across all classes, and the class with the highest probability is assigned as the predicted label. Each softmax node receives the feature map as an input and generates a probability distribution over the classes, as described in Equation (7). This process ensures that the classification head produces a reliable prediction, effectively distinguishing between different identities and capturing the relative confidence of the model for each possible class. Thus, the softmax layer serves as the final step in translating the learned feature representations into actionable classification outputs for biometric recognition.

$$p_j = \frac{e^{u_j}}{U}, j = 1,2,\ldots\ldots, C \quad (7)$$

Where, u_j represents weighted average calculated from FO, expressed as Equation (8):

$$u_j = \sum_{i=i}^{N}((w_{ij} * FO_i) + b_j) \quad (8)$$

w_{ij} is the weight vector connecting the dense layer nodes to j-th softmax node, with b_j as its bias. The computed value u_j is negative and not unitary, normalized using Equation (9), where U is computed as:

$$U = \sum_{j=1}^{c}(e^{u_i}) \quad (9)$$

p_j is expressed as vector of size $1 \times C$, given by Equation (10):

$$p_j = \frac{e^{(w_{ij} * FO_i)}}{\sum_{j-=1}^{C} e^{(w_{ij} * FO_i)}} \quad (10)$$

The predicted label is obtained from the value of j in vector p_j with the maximum value.

4 Result and Discussion

The experiments described in this section are conducted using the PYTHON to verify and validate the effectiveness of the multimodal biometric (MB) authentication system. The proposed CSMG method is evaluated on three benchmark datasets for improved classification. In the proposed CSMG method, local features were extracted using a CNN, while attention-based features were extracted from each biometric modality. After concatenation and normalization, features from all modalities are input to the classification head to produce the final result.

4.1 Evaluation Metrics

The evaluation metrics employed for assessing the performance of the multimodal biometric authentication system are presented in Table 1.

Table 1. Evaluation metrics to validate the performance of proposed CSMG.

Metric	Formula
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$
Precision	$\frac{TP}{TP + FP}$
Recall	$\frac{TP}{TP + FN}$
F1 Score	$2 \times \frac{precision \times Recall}{precision + Recall}$

4.2 Performance Analysis

The performance of the proposed CSMG method is evaluated on three benchmark datasets and compared with traditional methods such as SVM, RNN, and LSTM. In terms of accuracy, the proposed CSMG method achieved 99.90% on the SOCOfing fingerprint dataset, 99% on the IITD iris, and 99% on HEARTPRINT, significantly outperforming the traditional models. Similarly, the proposed CSMG method also demonstrated high recall, precision, and F1-score compared to SVM, RNN, and LSTM. Overall, the results demonstrated that the CSMG method is highly suitable for multimodal biometric classification. Table 2. presents an evaluation of the CSMG method performance.

Table 2. Performance analysis of proposed CSMG method.

Datasets	Models	Accuracy	Precision	Recall	F1-score
SOCOfing fingerprint	SVM	76	84	85	84
	RNN	76	86	83	77
	LSTM	77	79	76	81
	Proposed Method (CSMG)	99.90	98.94	99.50	99.87
IITD Iris	SVM	85	78	87	88
	RNN	87.	81	83	82
	LSTM	70	72	79	90
	Proposed Method (CSMG)	99	99.20	98.99	98.90
HEARTPRINT ECG	SVM	94	87	78	87
	RNN	94	86	86	85
	LSTM	94	90	81	81
	Proposed Method (CSMG)	99	98.93	98.97	98.96

4.3 Computational Complexity

The proposed CSMG method is evaluated for training and inference times compared with traditional methods such as SVM, RNN, and LSTM. The proposed CSMG method training time of 99 s is lower than that of SVM (133 s), RNN (154 s), and LSTM (123 s). Similarly, the inference time of the CSMG method is 14 as compared to SVM, RNN, and LSTM. The results of the CSMG method reduce the training and inference times, making it highly efficient. The proposed CSMG method is computationally lightweight and is suitable for multimodal biometric classification tasks, Table 3 shows the complexity of deep learning models.

Table 3. Complexity of deep learning models.

Models	Training Time (s)	Inference Time (s)
SVM	133	38
RNN	154	44
LSTM	123	31
Proposed (CSMG)	99	14

4.4 Statistical Analysis

Compared to the traditional CNN, Swin Transformer, and CNN+Swin+MHSA methods, the proposed CSMG method shows high accuracy and adaptability across diverse biometric modalities. The improvements in classification recall, accuracy, F1-score, and precision were statistically significant ($p < 0.05$), as confirmed by a paired t-test performed over 10 independent runs. This result confirms the reliability and effectiveness of the proposed CSMG method for secure and scalable biometric authentication.

4.5 Ablation Study

The ablation study in this section demonstrates that each component of the proposed CSMG method enhances performance using the SOCOfing fingerprint, IITD iris, and HEARTPRINT ECG datasets. Starting with CNN, the accuracy is comparatively lower, but adding the Swin Transformer improves the feature extraction and increases accuracy. When MHSA is included, the method becomes more effective at capturing global dependencies, leading to a better F1-score, precision, and recall. Finally, integrating Global Max Pooling strengthens feature selection and reduces redundancy, enabling the proposed CSMG method to achieve a high performance. Each added component contributes significantly, enhances the method, and achieves a high classification accuracy and reliability, Table 4 presents the ablation study conducted to evaluate the contribution of individual components in the proposed CSMG method.

Table 4. Ablation study of individual components in the proposed CSMG method.

Dataset	Models	Accuracy	Precision	Recall	F1-score
SOCOfing Fingerprint	CNN	94.20	91.85	90.72	91.20
	CNN + Swin Transformer	96.35	94.10	93.85	93.92
	CNN + Swin Transformer + MHSA	98.10	97.20	96.90	97.05
	Proposed (CSMG)	99.90	98.94	99.50	99.87
IITD Iris	CNN	92.10	89.50	90.00	89.75
	CNN + Swin Transformer	94.75	92.10	91.85	91.97
	CNN + Swin Transformer + MHSA	97.20	96.15	95.85	96.00
	Proposed (CSMG)	99	99.20	98.99	98.90
HEARTPRINT ECG	CNN	93.80	91.00	90.45	90.70
	CNN + Swin Transformer	95.90	93.75	93.40	93.57
	CNN + Swin Transformer + MHSA	97.85	96.30	96.00	96.15
	Proposed (CSMG)	99	98.93	98.97	98.96

4.6 Comparative Analysis

Table 5. The Comparative analysis of SOCOfing Fingerprint, IITD Iris and HEARTPRINT ECG datasets.

Datasets	Models	Accuracy
SOCOfing Fingerprint	Multimodal Biometric Recognition [11]	96
	LZS-GAN [14]	98.73
	Proposed Method (CSMG)	99.90
IITD Iris	Multimodal Biometric Recognition [11]	97
	CNN+GRU [15]	0.98
	Proposed Method (CSMG)	99
HEARTPRINT ECG	Multimodal Biometric Recognition [11]	71
	BXgBoost [12]	98.2
	GAM [13]	97.8
	Proposed Method (CSMG)	99

A comparative analysis of models on the IITD iris, SOCOfing fingerprint, and HEARTPRINT ECG datasets showed that the proposed CSMG method achieved the highest accuracy across all datasets. For the SOCOfing fingerprint, Multimodal Biometric Recognition reached 96%, KCP-DCNN was evaluated, and CSMG outperformed both by 99.90%. On the IITD iris, the multimodal baseline score 97%, and the CSMG improved performance to 99%. For HEARTPRINT ECG, Multimodal Biometric Recognition achieved 71%, BXgBoost 98.2%, and GAM 97.8%, while CSMG achieved 99%. These results demonstrate that the CSMG method provides superior performance in multimodal biometric classification, effectively combining fingerprint, iris, and ECG modalities, and surpasses the existing approaches in terms of accuracy and robustness, Table 5 Shows the comparative analysis.

5 Discussion

The experimental evaluation demonstrates that the proposed CSMG method significantly outperforms existing methods across all three datasets. The combination of CNN, Swin Transformer, MHSA, and Global Max Pooling enables the effective extraction of both local and global features, improving spatial and contextual representation. The feature fusion strategy successfully leverages the complementary strengths of fingerprint, iris, and ECG modalities, enhancing the recognition accuracy and robustness against spoofing attacks. The ablation study showed a positive contribution of each component to the overall performance, with Global Max Pooling reducing feature redundancy and MHSA capturing crucial global dependencies. Additionally, the CSMG method achieves superior computational efficiency with reduced training and inference times compared to traditional methods, making it suitable for biometric authentication applications.

6 Conclusion

This paper proposes the CSMG method, a multimodal biometric authentication system combining CNN, Swin Transformer, MHSA, and Global Max Pooling, to extract hierarchical and contextual features from fingerprint, iris, and ECG modalities. The CSMG method effectively fuses complementary features, improves the classification accuracy, and mitigates spoofing attacks. Experimental results on the SOCOfing fingerprint, IITD iris, and HEARTPRINT ECG datasets show that the CSMG method outperforms existing methods, achieving 99.90% accuracy for fingerprints, 99% for irises, and 99% for ECG, along with recall, high precision, and F1-score. Ablation studies confirmed the contribution of each component, while computational analysis demonstrated reduced training and inference times, making the method suitable for classification tasks. Overall, the CSMG method provided a reliable, efficient, and high-performance solution for secure multimodal biometric classification. Future work will focus on reducing the computational complexity, adding more biometric modalities, evaluating larger diverse datasets, and enhancing real-time deployment, adaptive learning, and spoofing resilience.

Reference

1. S.B. Abdullahi, Z.A. Bature, P. Chophuk, A. Muhammad, Sequence-wise multimodal biometric fingerprint and finger-vein recognition network (STMFPFV-Net). Intell. Syst. Appl. **19**, 200256 (2023). https://doi.org/10.1016/j.iswa.2023.200256
2. F.W. Ipeayeda, M.O. Oyediran, S.A. Ajagbe, J.O. Jooda, M.O. Adigun, Optimized gravitational search algorithm for feature fusion in a multimodal biometric system. Results Eng. **20**, 101572 (2023). https://doi.org/10.1016/j.rineng.2023.101572
3. S.P. Singh, S. Tiwari, A dual multimodal biometric authentication system based on WOA-ANN and SSA-DBN techniques. Sci. **5**, 10 (2023). https://doi.org/10.3390/sci5010010
4. M. Hammad, M.A. Wani, K.A. Shakil, H. Shaiba, A.A. Abd El-Latif, Deep cancelable multibiometric finger vein and fingerprint authentication with non-negative matrix factorization. IEEE Access. **12**, 120638-120660 (2024). https://doi.org/10.1109/ACCESS.2024.3450372
5. S. Artabaz, L. Sliman, Feature fusion and selection using handcrafted vs. deep learning methods for multimodal hand biometric recognition. Sci. Rep. **15**, 29237 (2025). https://doi.org/10.1038/s41598-025-10075-1
6. S. Salturk, N. Kahraman, Deep learning-powered multimodal biometric authentication: integrating dynamic signatures and facial data for enhanced online security. Neural Comput. Appl. **36**, 11311-11322 (2024). https://doi.org/10.1007/s00521-024-09690-2
7. V. Vekariya, M. Joshi, S. Dikshit, Multi-biometric fusion for enhanced human authentication in information security. Meas.: Sensors **31**, 100973 (2024). https://doi.org/10.1016/j.measen.2023.100973
8. S. Vatchala, C. Yogesh, Y. Govindarajan, M.K. Raja, V.P.A. Ganesan, A.A. Vinod, D. Ramesh, Multi-modal biometric authentication: Leveraging shared layer architectures for enhanced security. IEEE Access. **13**, 28029-28041 (2025). https://doi.org/10.1109/ACCESS.2025.3534223
9. T.S. Sasikala, Multimodal secure biometrics using attention efficient-net hash compression framework. Digit. Signal Process. **160**, 105018 (2025). https://doi.org/10.1016/j.dsp.2025.105018
10. T. L. Oteko, K.A. Ogudo, Dual Chaotic Diffusion Framework for Multimodal Biometric Security Using Qi Hyperchaotic System. Symmetry **17**, 1231 (2025). https://doi.org/10.3390/sym17081231
11. R. Garg, P. Pathak, M.P. Singh, A multimodal biometric recognition system based on fingerprints, iris and ECG via Swin Transformer and CNN model. Syst. Soft Comput. **7**, 200369 (2025). https://doi.org/10.1016/j.sasc.2025.200369
12. M.R. Prabhu, R. Sivaraman, N. Nagabhooshanam, R.S. Kumar, S.S. Salunkhe, Empowering artificial intelligence-based multi-biometric image sensor for human identification. Meas.: Sensors **33**, 101082 (2024). https://doi.org/10.1016/j.measen.2024.101082
13. F. Ahamed, F. Farid, B. Suleiman, Z. Jan, L.A. Wahsheh, S. Shahrestani, An intelligent multimodal biometric authentication model for personalised healthcare services. Future Internet **14**, 222 (2022). https://doi.org/10.3390/fi14080222
14. J. Vinayagam, G. Dilip, A two-step verification-based multimodal-biometric authentication system using KCP-DCNN and QR code generation. J. Ambient Intell. Humaniz. Comput. **15**, 3973-3996 (2024). https://doi.org/10.1007/s12652-024-04872-1
15. R. Rani, R. Dhir, K. Sonkar, Random projection-based cancelable iris biometrics for human identification using deep learning. Arab. J. Sci. Eng. **49**, 3815-3828 (2024). https://doi.org/10.1007/s13369-023-08190-0
16. SOCOfing fingerprint dataset link: https://www.kaggle.com/datasets/ruizgara/socofing (Accessed on August 18 2025)
17. IITD Iris dataset link: https://www.researchgate.net/figure/Sample-images-from-the-IITD-Iris-database_fig2_357404588 (Accessed on August 18 2025)
18. HEARTPRINT ECG dataset link: https://www.kaggle.com/datasets/shayanfazeli/heartbeat (Accessed on August 18 2025)